Autism Disclosures and Cybercrime Discourse on a Large Underground Forum

Jessica Man
Computer Science and Technology
University of Cambridge
Cambridge, United Kingdom
psjm3@cam.ac.uk

Gilberto Atondo Siu
Computer Science and Technology
University of Cambridge
Cambridge, United Kingdom
jga33@cam.ac.uk

Alice Hutchings
Computer Science and Technology
University of Cambridge
Cambridge, United Kingdom
alice.hutchings@cl.cam.ac.uk

***Abstract*—Prior research on the relationship between autism and cybercrime has been inconclusive. While some research suggests those who had an autism diagnosis are more likely to engage in cybercrime than those without, other evidence indicates a diagnosis of autism is associated with a lower risk of cybercrime offending. Prior research has primarily relied on self-report survey data. To the best of our knowledge, data gathered from cybercrime-related conversations on underground forums has not previously been used to study the relationship between autism and cybercrime offending. This research applies natural language processing (NLP) techniques to a large underground cybercrime forum data. We developed two NLP classifiers to automatically categorise the context and content of forum posts. We find that terms related to autism were mostly used in a negative context, primarily to insult other users. We find that actors who self-declare as autistic, post more frequently on the forum than those who do not disclose to be autistic. Despite the increased frequency of their activity, we find those who disclose they are autistic are less likely to discuss cybercrime-related matters, compared to a matched sample of users with similar posting activity.**

I. Introduction and Background

It is often assumed that there is a relationship between autism and cybercrime. Indeed, there are several alleged cybercrime offenders who have been extensively discussed in the media in relation to their autism diagnoses, including Julian Assange [1], Gary McKinnon [2], and Laurie Love [3]. Autism is a spectrum condition, affecting more than 1% of the UK population, according to the latest figures from the National Autistic Society [4]. There are other diagnostic labels used such as Autism Spectrum Disorder (ASD), Autism Spectrum Condition, Pervasive Development Disorder, High-Functioning Autism, and Pathological Demand Avoidance. In this paper, unless specified otherwise, all these terms are used interchangeably to mean autism spectrum conditions. The Autism Spectrum Quotient (AQ) is a self-administered instrument consisting of a 50 questions questionnaire (also known as AQ-50), invented by Baron-Cohen and published in 2001 [5], to quantify the degree to which an adult with normal intelligence has the traits associated with the autistic spectrum. AQ is widely used for assessing autistic traits but its effectiveness to predict a clinical diagnosis of ASD remains questionable [6], [7].

There is a general belief that there is an association between autism and an advanced level of technical skills. Baron-Cohen et al. [8] provided evidence that hyper-systemising and excellent attention to detail are part of the cognitive style of people with ASD, which means that there is a higher chance that they acquire advanced computer skills. Due to the technical nature of some types of cybercrime, the relationship between autism and cybercrime has been of research interest. Yet, the evidence from academic studies has been mixed [9], [10]. While popular media portrays cybercrime offenders as being autistic, some neurodiverse people have pushed back against the stereotype [11], [12].

The inconsistencies and inconclusive results from previous research inspired this work. In the context of cybercriminal communities, this research examines if there is a relationship between people's social behaviour and whether or not individuals self-disclose as being autistic. We build upon what has been studied in the past (e.g. [10], [13], [9]), moving beyond self-reported survey data. Our research uses data that brings us closer to where conversations and interactions between cybercrime actors take place – within underground forums. Underground forums are public online platforms where users exchange information and knowledge, express opinions, and trade tools and services. Forums can bring together individuals who are interested in cybercrime and illicit monetising techniques [14].

This research addresses the following questions:

1) What is the nature of autism-related conversations on a large cybercrime forum?
2) Are those who self-disclose to be autistic more or less likely to discuss cybercrime compared to a matched sample on the same forum?

To answer these questions we use the data provided by CrimeBB [15], a database containing data scraped from multiple online forums. Specifically, we select data from HackForums, the largest and most active underground forum, for this work. We make the following contributions:

- We build two classifiers that automate the labelling of underground forum posts according to their context and content related to autism. The autism-context classifier identifies the different autism-related context the conversations were in (e.g. self-claimed autistic or using the term to reference other actors or things) and the autism-

content classifier identifies the different topics related to autism in each post (e.g. cybercrime or general-health).
- We find that autism-related terms are mostly used as an insult to attack other forum actors, indicating that autism is not perceived positively amongst users.
- We encounter significant differences between autistic-claimed actors and non-autistic-claimed actors. First, actors who claim to be autistic are less likely to participate in cybercrime-related conversations on the forum. Second, for those who did talk about cybercrime, there was a significant difference in the types of crime being discussed.

Our work is organised as follows. We outline the mixed results from prior work investigating the relationship between cybercrime offending and autism in §II, along with an overview of research into underground forums. In §III we present our methods, including ethical considerations and classifier development. In §IV we evaluate five classifiers, finding XGBoost outperforms SVM, Logistic Regression, Random Forest and LSTM (with GloVe). This section also analyses the results of the content automatically labelled using XGBoost. For the crime type content, we use a text classifier previously developed by Atondo Siu et al. [16]. Our conclusion (§V) includes a discussion of the limitations and potential future research directions.

II. Related Work

A. Autism and Cybercrime

Ledingham and Mills [17] studied international law enforcement specifically in the context of autism and cybercrime. In 2015 they published the results from the survey on national policing agencies in multiple countries. They acknowledged that, whilst there is a growing awareness of mental health issues generally, where autism is specifically concerned, the data is incomplete. The systematic examination or recording of key data is lacking, as a result the assessment is frequently left to chance or dependent on local factors. The conclusion from this survey was that even though autistic individuals are known to have been involved in cybercrime, no estimate of prevalence could be made.

In the same year, Seigfried-Spellar et al. [13] published their work on assessing if autistic traits were significantly related to cybercrime. They invited 296 university students to complete an online anonymous survey that was designed to measure self-reported computer deviant behaviour and autistic traits, and found no evidence to support that clinical levels of autism has a significant relationship with self-reported computer deviance. However, those who engaged in computer deviancy did report more autistic-like traits.

In 2017, the UK's National Crime Agency (NCA) published a report on 'Pathways Into Cyber Crime' [18]. One of their key claims was that ASD appears to be more prevalent among cyber-criminals, but this remains contested. Their findings were based on interviews conducted by their officers with people involved in cyber-dependent cases. Opinions about ASD were largely subjective and the report did not specify how the anecdotal evidence was formed. They did point out that the evidence was not sufficient to infer any link, and that a more in-depth study was underway.

In 2019, Payne et al. [10] conducted an anonymous online survey with 290 participants (all with a computer science background) to find if autism, amongst other variables, was a predictor of cyber-dependent crimes. They found 8% of participants self-reported a diagnosis of autism and also had higher levels of autistic-like traits than those who did not have a self-reported diagnosis. Of the 122 participants who reported having committed cyber-dependent crimes, while they had higher AQ scores, and greater basic and advanced digital skills, they were less likely to report being diagnosed as autistic.

Payne et al. [19] did another study to look further into the self-reported motivations behind cyber-dependent offending, and how they relate to the level of autistic traits. They invited 175 cyber-skilled non-offenders and 7 cyber-dependent offenders to complete an online survey designed to measure participants' autistic traits and characteristics. 29 out of the 175 non-offenders reported that they had been approached to commit a cyber-dependent crime but had declined, this group of 'cyber-dependent decliners' (CDD) and the cyber-dependent offenders (CDO) were questioned further to find out their motivation for engaging or declining in offending. Only one participant self-reported a diagnosis of autism, and therefore cannot be used in any implications grading autism.

In 2021, Lim et al. [9] conducted a bigger survey, inviting 742 participants online and through contacts from a university autism database, 440 responses were rejected based on their priori exclusion criteria, which include a set of bogus questions. Their results found 25 out of the 302 participants reported to have a diagnosis of autism. They had higher scores on the AQ-12 (a 12-item version of AQ designed by Lundqvist and Lindner [6]) than the non-autistic participants and scored lower on the theory of mind test, although the difference was not statistically significant. The researchers identified that 36 participants reported having committed cyber-dependent crimes as measured by Cyber-Dependent Crime Questionnaire (CDCQ, developed by Payne et al. [10]), whilst 98 reported having committed cybercrimes as measured by Computer Crime Index-Revised (CCI-R [20]). Their analysis of CDCQ and CCI-R results found that autistic traits did not have a significant relationship with the scores on CDCQ or CCI-R. However, autism diagnosis did significantly predict cybercrime as measured by CCI-R. Hence this conclusion went in a different direction from Payne et al.'s findings.

In 2022, the NCA published another report [21] based on those referred for Cyber Prevent or Pursue activity between 2017 and 2020 due to suspected offending. They reported that the rate of people who either had a diagnosis for ASD or self-disclosed as having autistic traits (17%) was far higher than ASD diagnosis in the general population (1-2%). This claim might be misleading as the 17% included both diagnosed autistic individuals and those self-reported, and there was no clear indication of how many of the 17% were not-diagnosed.

While Payne et al. [10] surveyed individuals with an education in computer science, Lim et al. [9] recruited participants from the general public and a university autism database. Our research takes a different approach, instead of using self-reporting data, we use data directly from HackForums. Our results support Ledingham and Mills [17] and Payne et al.'s [10] research, finding that self-claimed autistic people are less likely to be involved in cybercrime-related activities.

B. Underground Forums and CrimeBB

Underground forums provide a platform for cybercrime offenders to participate in illicit activities. Forums provide ways for users to share values, attitudes, techniques, and motives for criminal behaviour [22], which are theorised to be important for crime commission [23]. These forums bring together people who have an interest in cybercrime, not only to share information and services but also to trade and monetise their techniques [15], [24], [25]. Therefore, data from these forums is useful to learn about criminal activities as well as how actors interact.

For our research, we use the CrimeBB dataset, which contains scraped data from cybercrime forums [15]. The dataset is made available for academic research through data sharing agreements with the Cambridge Cybercrime Centre.[1] At the time of writing, CrimeBB contains more than 100 million posts, some dating back more than 20 years. The dataset has been extensively used for cybercrime research (e.g., [14], [16], [24], [26], [27], [28], [29]).

In this research, we applied Sykes and Matza's Techniques of Neutralization theory [30]. Sykes and Matza outlined how juvenile delinquents use techniques to excuse or justify their actions: denial of responsibility, denial of injury, denial of the victim, condemnation of the condemners, and appeal to higher loyalties. Prior research has found evidence of these neutralisations being used in relation to cybercrime offending [31], [32], [33], [34]. In the classification task to identify the topic of discussion, we look to see if actors are using autism as a way to deny responsibility for taking part in cybercrime-related activities.

C. Underground Forums and Data Analysis

Given the large volume of data collected from underground forums, manual processing would be hugely inefficient. Also, conversations on these forums tend to use slang and jargon, it is a challenge for standard NLP tools to perform well provided the use of non-standard language, and domain expertise requirements. There is a growing body of work developing tools for analysing these messy datasets. Portnoff et al. [35] developed NLP models for data extraction tasks. They applied the techniques to two case studies. The first case was to identify account activity. They found that plural headwords (e.g. accounts, emails) almost always reflect illegally acquired accounts trafficking, whereas singular headwords reflect users selling their accounts. Their second case, identifying currency exchange patterns, found the most popular exchange offered was Bitcoin for PayPal. This demonstrated the applicability of the tools for large-scale automated exploration on underground forums. They have also discussed the limitations due to the non-grammatical language used and differences between the forums.

Caines et al. [26] also developed NLP tools for data extraction tasks from underground forums using the CrimeBB corpus. Their tasks were to classify the function and intent of texts to identify the key actors, tools, and techniques in conversations. They applied a similar approach to Portnoff et al. [35], by first manually labelling a sample from the corpus, using it as the training set for supervised training, and then using the classifier to label posts in CrimeBB for further analysis. They tested logical models, statistical models and linear models, and concluded that the best is a hybrid of logical and statistical for post type and author intent.

Similarly, Pastrana et al. [14] applied NLP techniques to HackForums data to extract information related to cybercrime. Their focus was on identifying and predicting the key actors on HackForums based on their activity and social relations. Their method involved applying social network analysis to build a network of key actors based on the type of their social interactions (positive, negative, and neutral/unknown), and machine learning algorithms (k-means clustering) to characterise key actors. Their results showed that most key actors were closely connected and their social relationships were mainly positive. Their cluster analysis suggested that key actors are mostly interested in market, common and hacking areas, and over time their interests on coding and technology increased slightly, whilst their interests on gaming dropped.

Atondo Siu et al. [16] used HackForums data to seek the links between illicit behaviour and currency exchange. They developed a classifier that categorises HackForums posts by crime type. They found that the most popular topics of discussion were related to trading credentials and bots/malware when exchanging currencies was involved. They evaluated four statistical models and found XGBoost to perform the best. The forum data classified by crime type was used in this work for the analysis of cybercrime-related activities, so we could compare autistic-claimed and non-autistic-claimed actors on HackForums.

This research found a similar result to Atondo Siu et al.'s [16] work, that XGBoost out-performed all the other tested models. Two classifiers were developed using similar NLP techniques as the research mentioned above: an autism-context classifier to extract posts in autism context, and an autism-content classifier for categorising the content of conversations.

III. Research Method

A. Dataset

The dataset was extracted from CrimeBB [15] provided by the Cambridge Cybercrime Centre [36]. The full dataset consists of posts from HackForums specifically, in total there were 42,112,205 posts across all boards on HackForums, by

[1] https://www.cambridgecybercrime.uk/process.html

640,458 unique actors, dated from March 3, 2008 to August 8, 2020.

B. Ethical Considerations

This research was approved by the department's research ethics committee. Users who posted on HackForums would have agreed to the privacy policy set by the forum platform, which states that when signing up the user agrees they do not expose any information that would identify them or another person. The data published via this platform is already publicly available on the internet, and the dataset extracted is used for research without aiming to identify any particular individual. To ask for consent from all the forum users would not be feasible. Any data referenced in this report is shown anonymously, and all extracted text has been paraphrased.

C. Data Extraction

According to the National Autistic Society [4], the definition and label of autism has changed over the decades and could continue to change in future years. The terms used also depend on the person, as for some people this forms a core part of their identity. On HackForums however, yet more different terms could be used, including when used to reference other actors. On HackForums, we looked at some sample posts and found the following terms frequently used to mean the autism spectrum conditions:

autism, autistic, autist, asperger(s), asberger(s), aspie(s)

These terms were used in an SQL query to extract all posts from CrimeBB that contain one (or more) terms listed above. Also extracted were post IDs, IDs of the actors who posted them, and board names they were posted in (boards are sub-forums within HackForums for actors to post specific topics). There are other autism-associated terms such as ASD, ASC, AQ and 'the spectrum', but they mean different things on the forum and hence were not included in this data extraction query.

This query returned 19,849 posts from 6,887 unique actors. This dataset was used as our gold standard corpus and has two main purposes: first, to find out the nature of the conversation when autism was used in context, this could be resolved by categorising each post based on the text used, as described below in §III-D. The second purpose was to find the actors who self-disclosed that they were autistic, and from that we have the autistic-claimed population vs the non-autistic-claimed population for the wider analysis across all posts within the forum (§III-G).

D. Manual Annotation

As we used supervised machine learning approaches, we labelled randomly selected samples of the extracted data for training our classifiers. Supervised machine learning involves analysing labelled data and training the algorithm until it can detect patterns and relationships (also known as predictor features) in the data and labels [37]. Three annotators manually labelled randomly selected posts according to their context and content.

The autism-context classification relates to the context in which the actor uses the autism-related terms. For those that identified as being autistic (e.g. 'I am autistic'), we initially annotated the data to differentiate between those who revealed they had a diagnosis, those that indicated they did not have a diagnosis, and those that did not say either way. Autism-referencing was used when the author referred to autism in their posts, but in relation to others rather than themselves (e.g., 'This is the most autistic thing I've seen'). The annotation guidelines for context are shown in Table I, which provide a description of each category and some anonymised examples.

The autism-content classification relates to the topic of discussion in the context of autism. The annotation guidelines are provided in Table II. The neutralisation category relates to autism being used as an excuse for offending, which may be thought of as denying responsibility, one of the neutralisations proposed by Sykes and Matza [30], as discussed in §II-B. Other categories included cybercrime more generally, but also general mental health, or topics that portray autism in a positive, negative, or neutral way.

The annotation task was completed over four iterations, with each iteration carried out by the same annotators. We encountered a number of challenges during this task. The first challenge was that as each post is a snapshot of a conversation, often there was not enough information in the post itself to judge the intention. For example, "hf [HackForums] autism still here, always will be" could be negative or positive. There were posts containing only the word 'Autism', which lacked a full sentence. As we progressed with annotations, finding that autism was used in a derogatory way on the platform, we classified these as 'negative'. The second challenge was the non-standard language used, which includes sarcasm. Sentences could be ambiguous, or have a completely different meaning if taken out of context. For example, "Autism at it's finest" has a positive word in the sentence, but this could be a sarcastic comment. The third challenge was the lack of understanding of the language used on the forum. There were acronyms and slang terms that without the knowledge of their meanings would be difficult to understand. We overcame these challenges by holding regular moderation meetings to discuss disagreements in our annotations and came to a consensus.

Initially, when we had the six independent classes for the content classification there was a mix of interpretation between cybercrime and negative-connotation, and between general-health and neutral-connotation. We decided to use a hierarchical system to make it clearer for the annotators to differentiate between them – the classes 'Neutralisation', 'Cybercrime' and 'General-Health' were at the top-level. Anything that did not fall into any of these classes was then classified as negative, positive, or neutral-connotation. This influenced the choice of using a hierarchical classifier as one of the models we evaluated, but this was rejected due to poor performance.

We annotated four randomly selected samples. The first sample, containing 317 posts, was selected from all posts containing autism-related terms. After our initial annotations,

TABLE I
AUTISM-RELATED CONTEXT ANNOTATION GUIDELINES

Category	Description	Anonymised example
Claimed-Diagnosed	Users who have stated that they have been medically diagnosed or assessed. For example, they might express that they have had a positive autism assessment, or are receiving treatment (psychological, social or pharmaceutical).	"Basically I have been diagnosed with Asperger's. My family are devistated, just because."
Self-Diagnosed	The main difference between a claimed-diagnosed user and a self-diagnosed user is whether or not the user has claimed a completed diagnosis. A self-diagnosed user would say something like "I think I am autistic", "I must be autistic because..." or "I have high scores on the AQ questionnaire" rather than more certain phrases such as "I have been diagnosed as autistic".	"I know I have Aspergers Syndrome, I've been self diagnosing myself lately and that's the conclusion I have"
Claimed	Users who fall into this category have a strong claim but without further information to support it. They would say that they have autism or they are autistic but provide no diagnosis claims.	"Right ok so I have autism. Does anyone else here has it or aspergers too. Could you please clarify if you do?"
Denied	Reject or decline claims that they are autistic, or self-claim that they do not have autism.	"me too, I'm not at all autistic. It's not that uncommon."
Autism-Referencing	These users referenced autism in their posts. They do not necessarily associate the term with themselves but might quote or mention autism to support the points they make on the posts.	"I know lots of autistic hackers, they do nothing else xD or learn nothing else, when they concentrate on RATs"

TABLE II
AUTISM-RELATED CONTENT ANNOTATION GUIDELINES

Category	Description	Anonymised example
Neutralisation	Autism being discussed as an excuse for taking part in cyber-criminal activities (denial of responsibility).	"2 charges, hacking and extortion. I was on bail for 2 years. In the end I was found not guilty though, because I am autistic. You know."
Cybercrime	The posts discuss cybecrime in autism context, either committed by the author or another party. This includes the mentioning of relationship between autism and cybercrime.	"Someone get me the admin password, I'm going to unban all accounts on the forum. It's a psychforum where i was targed for having Autism by the stupid admin."
General Mental Health Topic	Discussions on being autistic, the effects on them, the autistic traits, the reasons why the interest on the topic, and treatments.	"basically a form of high-functioning autism. I don't like going out or speak to new people, I get pissed off when my routine is being disrupted. I'm very hardcore about video games, computers, malware in general, collecting random and games stuff"
Positive Connotation	Autistic traits are being valued and praised.	"Since most people with asperger got a IQ higher then the average person he/she will likely understand coding faster then the average person."
Negative Connotation	The autism related terms were used as an insult or attack to other actors in the conversation.	"I was reading the thread to see the amount of autism in responses to this question, but you actually made me respond to this cancerous thread, you're so much of an idiot."
Neutral Connotation	Not specifically being positive or negative regarding autism.	"I am hoping to move to a better University, could someone help me with a bit of advice/proof reading? I do have aspergers."

we found the autism-context data was largely imbalanced (one of the classifications 'autism-referencing' dominated, at over 86% posts). In the second round, we selected 303 posts from a reduced dataset that removed boards used mainly for social interactions (e.g. 'The Lounge'). While annotations remained imbalanced, inter-related agreement improved, the agreement on autism-context annotation went up by 6% using Fleiss's κ measurement (see next paragraph below) and on autism-content annotation it went up by 28%. For the third round, we randomly selected 400 posts from boards selected due to the likelihood they would contain cybercrime-related conversations. The difference in agreement between the annotators was greater than all previous rounds, so we did one final round of 100 posts. For this round we made use of the data annotated using Atondo Siu et al.'s [16] crime type classifier, and selected from posts predicted to relate to criminal activity.

As there were three annotators, we used Fleiss's κ to quantify the degree of agreement between the annotators. Fleiss's κ is a statistical measure on nominal scale agreement between multiple raters giving categorical ratings (labels in our case) [38]. $\kappa = 1$ means all annotators are in complete agreement, while $\kappa \leq 0$ means the annotators disagreed totally (other than what is expected by chance). The measurements for the four iterations on each round of manual annotation are shown by Table III. This table also shows Landis and Koch's [39] interpretation, where 'Poor' and 'Slight' are when $\kappa \leq 0.20$, 'Fair' is when $0.20 < \kappa \leq 0.40$, 'Moderate' is when $0.40 < \kappa \leq 0.60$, 'Substantial' is when $0.60 < \kappa \leq 0.80$, and

'Almost Perfect' is when $\kappa > 0.80$.

In total we extracted 1,120 posts for the annotation exercise, but found some posts were not related to the autism condition, for example a name reference could contain the spelling of our search terms. Having rejected these non-relevant posts we ended up with 1,116 annotated posts as the training dataset.

E. Automated Data Classification

Methods and techniques for NLP have advanced rapidly in recent years, with each approach designed for specific objectives. How well they perform depends on the data structure, text characteristics, and the relationship between the data and the classifications. We test and evaluate to select the most suitable model, but first, we need to narrow down the list of models to test. The selected models are the most popular models for text classification based on past research on cybercrime forums: **Support Vector Machines (SVM)** [16], [26], [14], [35]; **Logistic Regression** [16], [26]; **Random Forest** [16] and **eXtreme Gradient Boosting (XGBoost)** [16], [26]. We also experimented with **Long Short-Term Memory (LSTM)** [40], to find out if a neural network approach would be beneficial in our text classification. Typically LSTM is used for patterns and time series prediction, but it has been applied to text classification and achieved good results [41], [42].

We evaluate the classification models in progressive steps. Using SVM and Logistic Regression as the baseline, we subsequently test the selected models, i.e. Random Forest, XGBoost, LSTM, with the aim to improve baseline scores through experimenting with different parameters and tuning. The performance measurements we use for the models evaluations are **Precision**, **Recall**, **Accuracy** and **F-measure**, metrics commonly used to quantify the performance of classification and other information retrieval tasks.

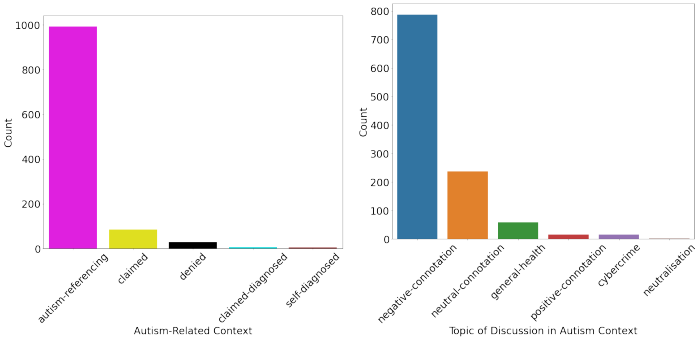
We apply a number of pre-processing steps to convert the raw strings of posts into a format that is suitable for the classifiers, such as tokenisation (breaking raw text and sentences into word tokens), lemmatisation (converting words into their base form based on their intended meaning) and word embedding (converting word tokens into a matrix filled with the words' occurrence counts within each post, and subsequently transformed into a weighted vector) [43]. We apply the **'Term Frequency - Inverse Document Frequency' (TD-IDF)** algorithm to this processing. TD-IDF uses the weighting which reflects how important a word is within the corpus, based on not only the number of times the word appears in a post but also on the number of posts within the full corpus containing that word. There are other techniques such as **'GloVe', Global Vectors for Word Representation**, which is typically used in unsupervised learning. We apply GloVe in our LSTM model evaluation.

Before the dataset was ready for the evaluation of different classifier models, there were two issues to tackle: imbalanced data and overfitting. As mentioned above, we noticed the annotated dataset was largely imbalanced, with the majority of autism-context classification resulting in 'autism-referencing' and the majority of autism-content classification resulting in 'negative-connotation'. Overfitting is when the trained model over-optimises on one set of data, such that it can only do well with data it has seen already but not so much on unseen data. Imbalanced input datasets and overfitting tend to generate unwanted biases in NLP. A common technique to tackle imbalanced datasets in NLP is re-sampling the data by down-sampling the majority classes or over-sampling the minority classes. We use the Imbalanced-learn Python API [44] to do the re-sampling. We considered two techniques: The random over-sampling technique that uses an algorithm to generate new samples in the classes that are under-represented, and the Synthetic Minority Over-sampling Technique (SMOTE) [45] that uses synthesized samples from the existing data and adds to the results dataset. Experiments showed that up-sampling with SMOTE performed the best, and hence was chosen to be part of the autism-context classifier to predict the full dataset.

To optimise the performance of the models during evaluation, a technique called hyper-parameter tuning was used. A hyper-parameter is external to the model used during training and cannot be determined using the data, such as the learning rate at which the algorithm updates the estimates, or the number of hidden layers in a neural network. Different parameters and values were tested to determine the combination that achieved the best performance, using techniques Cross-Validation and Grid Search. Cross-Validation splits the data into K subsets called folds. The algorithm iteratively train the model K times, each time train on $K-1$ of the folds and evaluate on the Kth fold. At the end of the cycles the results are compared and the best combination is returned. Grid search algorithm exhaustively searches through the hyper-parameters set to do the training. Typically bounded values are specified for the algorithm, such that it can perform the search finitely.

F. Reducing Classes by Merging Categories

Recall that we have 5 autism-context and 6 autism-content categories. Out of the 1,116 posts annotated, 89% were labelled 'autism-referencing' in the autism-context classification and 71% were labelled 'negative-connotation' in the autism-content classification. Figures 1a and 1b show how the categories are distributed in the training dataset.



(a) Autism-context categories (b) Autism-content categories

Fig. 1. Categories distribution in training dataset

TABLE III
FLEISS'S κ SHOWING DEGREE OF AGREEMENT FOR AUTISM-CONTEXT AND AUTISM-CONTENT ANNOTATIONS

		Autism-context		Autism-content	
Iterations	N	κ	Agreement	κ	Agreement
1 - All extracted HackForums data	317	0.84	Almost Perfect	0.43	Moderate
2 - Excluding non-cybercrime-related boards	303	0.89	Almost Perfect	0.55	Moderate
3 - Including only cybercrime-related boards	400	0.73	Substantial	0.45	Moderate
4 - Including only crime-type-classified data	100	0.85	Almost Perfect	0.63	Substantial

As that the end goal for this stage from the autism-context classification is to obtain the list of autistic-claimed actors on HackForums, we interpret the classes 'claimed', 'self-diagnosed', and 'claimed-diagnosed' all as actors who have self-disclosed as autistic. Therefore, we collapse these three categories into a single 'claimed' class. Likewise, as 'autism-referencing' and 'denied' can be interpreted as autism not being in the context of self-disclosing, we combine them into a single 'autism-referencing' class. This also simplifies the task into binary classification. For the autism-content classification, we collapse the 'neutralisation' and 'cybercrime' categories. We note there are only two posts manually annotated as 'neutralisation'.

G. Analysis with a Matched Sample

To answer the second research questions, *Are those who self-disclose to be autistic more or less likely to discuss cybercrime compared to a matched sample on the same forum?*, we use an analytical approach. For the statistical test we need two samples, one for the autistic-claimed population and one for the non-autistic-claimed population. Due to the high posting volumes for the autistic-claimed sample (see §IV-D), we obtain a matched sample based on posting volume. The post counts **c** are divided into count ranges with 500 posts in each range, i.e. $[0 < \mathbf{c} \leq 500]$, $[500 < \mathbf{c} \leq 1000]$ and so on up to 20,000 posts. Two actors on the forum posted over 20,000 times (one of them posted over 85,000). We exclude these outliers to reduce the likelihood they have an undue influence on results, which uses posts as the unit of analysis. Once we have the number of autistic-claimed actors in each range, we extract the same number in each range for the non-autistic-claimed population to obtain a matched sample. We then obtain the predicted crime type of all the posts in the samples, by first running a query to CrimeBB to get all the IDs of the posts by those actors and running a cross-referenced query to the database that already has the posts classified by crime types [16]. The crime type classification used by Atondo Siu et al had an imbalanced data set with the majority of posts classified as "Not Criminal", however the model they have chosen achieved a high percentage (over 87%) without over-fitting, which gave us enough confidence to use it. Our data set came from the same HackForums and therefore we expected to see similar classification results.

We use Pearson's χ^2 test of independence [46] to test if there is a significant difference between these two populations. Due to the mixed nature of prior research, we do not hypothesise the direction of any relationship, such as posts by autistic-claimed actors containing more or less cybercrime-related material than those by non-autistic-claimed actors. Instead, we take an exploratory approach, carrying out two statistical tests using post volume. The first χ^2 test compares posts classified as 'Criminal vs Not Criminal', using the combined number of post count of all crime types and the post count of the 'not criminal' type. The second χ^2 test for 'Criminal Types', uses the full set of crime type classified post count, excluding the 'not criminal' type.

We estimated the χ^2 value using the equation:

$$\chi^2 = \sum \frac{(O-E)^2}{E} \tag{1}$$

where *O* is the observed frequency of posts in each crime type by each actor, and *E* is the expected frequency of posts in each crime type by each actor, calculated using equation 2.

$$\frac{T_{crime} \times T_{actor}}{T_{all}} \tag{2}$$

where T_{crime} is the total of counts for each crime type, T_{actor} is the total of counts for each actor type, and T_{all} is the total counts for all.

We also calculated the standardised residual, given by equation 3, to show the strength of the differences between the observed and expected for each value. A cell in the χ^2 grid that has an absolute residual value bigger than 2 means that it has a significant deviation from the expected value and hence a strong contributor to the χ^2 result.

$$c = \frac{(O-E)}{\sqrt{E}} \tag{3}$$

IV. RESULTS AND ANALYSIS

A. Model Evaluation Results

Between the baseline models, Logistic Regression performed better overall in the autism-context classification. Whilst for autism-content classification SVM did better across all the scores. After applying the up-sampling technique and with the classes reduced to two, there were true positives in the 'claimed' class but there were also many false positives and false negatives, as indicated by the confusion matrix shown in Figures 2 and 3. The Logistic Regression model did not perform well at predicting the 'claimed' class, when it was used to predict the unseen data (i.e. the full dataset that was not pre-labelled) it returned zero 'claimed' results.

With the mission to improve on the performance from the baseline models, the non-linear models described in §III-E

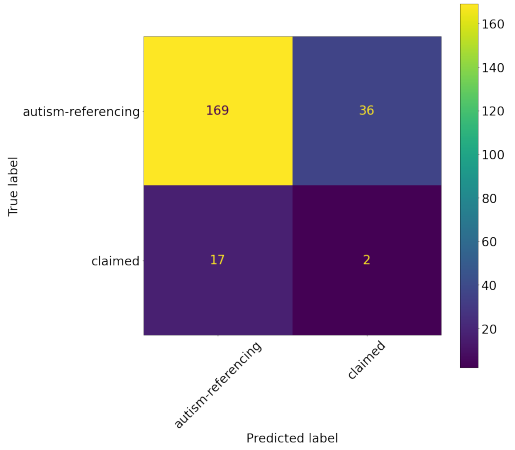


Fig. 2. Confusion matrix for autism-context classification with SVM

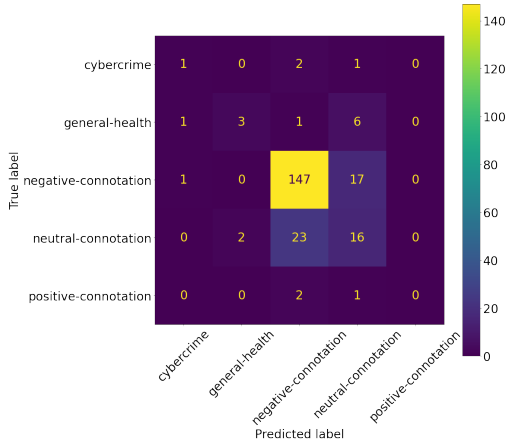


Fig. 3. Confusion matrix for autism-content classification with SVM

were experimented using the hyper-parameter tuning technique. This set of models all perform better than the baseline, with Random Forest and XGBoost performing the best on both autism-context and autism-content classification tasks. Results are shown in Tables IV and V.

B. Chosen Classifier and Automated Prediction

We selected XGBoost for both classifiers. For the autism-context classifier, compared to Random Forest, it has a slightly higher true positive count on the 'claimed' class and the model does better on the task using the original full set of classes. Similarly for the autism-content classification, direct comparison between Random Forest and XGBoost show that XGBoost performs slightly better. As for LSTM, the results are consistently poorer compared with the other models, because of the higher numbers of false positives and false negatives in its predictions. Table VI shows the full set of performance results from the training for the chosen model for the prediction.

Out of the 19,849 posts, 401 are classified as 'claimed' and 19,448 are classified as 'autism-referencing' in autism-context. For the autism-content classification, 24 are classified as 'cybercrime', 95 as 'general-health', 18,477 as 'negative-connotation', 1,246 as 'neutral-connotation' and 7 as 'positive-connotation'. To validate further we perform a manual check on a small sample of 153 posts from the 'claimed' classified set. 102 are correctly classified (0.67 precision). It gives us enough confidence to use the dataset, giving us 282 unique actors in the 'claimed' classified set. We increase this number by looking into the 'autism-referencing' classified set for false negatives. We search the phrases "I have Aspergers", "I am autistic", and "I'm autistic" (case insensitive) specifically from this set, validating them, which provides a further 102 unique 'autistic-claimed' actors to add to the population.

C. Autism-related Topic of Discussion Analysis

To answer the first research question, "What is the nature of autism-related conversations on a large cybercrime forum?", we analyse the autism-content classified data. The majority of the classified posts are in the 'negative-connotation category', with only a few classified as 'cybercrime'. This indicates the nature of the topic of discussion relating to autism tends to have a negative context. During our manual annotation exercise we already noticed this trend, with the term commonly being used as an insult. Despite the number of negative-connotation posts being much higher than the rest, the number of them that were also classified as 'claimed' by the autism-context classifier is small. The category 'neutral-connotation' has a higher number within this population. The distribution is shown in Figure 4, which shows a clear difference in the proportion of each category amongst the 'claimed' and 'autism-referencing' populations.

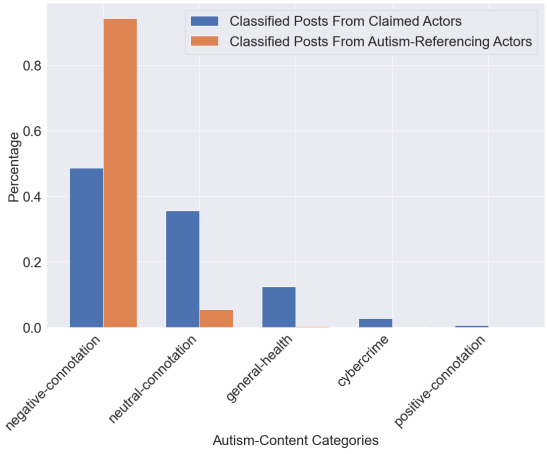


Fig. 4. Distribution of classified autism-content, with each class further separated in 'Claimed' and 'Autism-Referencing' classes

There are 11 posts classified in the 'Cybercrime' category out of the 401 claimed posts. Two are mis-classified as 'claimed' (one expresses their opinion negatively regarding how actors post on 'The Lounge' board on HackForums and implicitly calls another actor autistic, the other one wrote a story about an autistic hacker). Of the remaining nine, one should be classified as 'positive-connotation' because the actor talks about how being autistic makes them "extremely intelligent", and another one should be in 'general-health' as the

TABLE IV
CLASSIFIER PERFORMANCE METRICS FOR AUTISM-CONTEXT CLASSIFICATION

	Original Classes				Up-sampled Reduced Classes			
Classifier	Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy
SVM	0.87	0.85	0.78	0.85	0.85	0.81	0.83	0.81
Logistic Regression	0.84	0.89	0.85	0.89	0.89	0.93	0.91	0.93
Random Forest	0.91	0.90	0.85	0.90	0.90	0.92	0.91	0.92
XGBoost	0.87	0.90	0.87	0.90	**0.90**	**0.92**	**0.91**	**0.92**
LSTM (with GloVe)	0.84	0.89	0.86	0.89	0.89	0.87	0.88	0.87

TABLE V
CLASSIFIER PERFORMANCE METRICS FOR AUTISM-CONTENT CLASSIFICATION

	Original Classes				Up-sampled Reduced Classes			
Classifier	Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy
SVM	0.79	0.69	0.56	0.69	0.70	0.71	0.69	0.71
Logistic Regression	0.68	0.71	0.67	0.71	0.64	0.69	0.66	0.69
Random Forest	0.67	0.69	0.63	0.69	0.67	0.67	0.60	0.67
XGBoost	0.66	0.68	0.61	0.68	**0.66**	**0.70**	**0.68**	**0.70**
LSTM (with GloVe)	0.67	0.71	0.67	0.71	0.68	0.63	0.65	0.63

TABLE VI
PERFORMANCE METRICS FOR THE CHOSEN AUTISM-CONTEXT AND AUTISM-CONTENT CLASSIFIER

	Precision	Recall	F1	Accuracy
(context) autism-referencing	0.94	1.00	0.96	0.93
(context) claimed	0.75	0.18	0.29	0.93
(context) weighted avg	0.92	0.93	0.91	0.93
(content) cybercrime	0.00	0.00	0.00	0.93
(content) general-health	0.00	0.00	0.00	0.93
(content) negative-connotation	0.79	0.97	0.87	0.75
(content) neutral-connotation	0.40	0.14	0.20	0.75
(content) positive-connotation	1.00	0.00	0.00	0.75
(content) weighted avg	0.67	0.75	0.68	0.75

actor describes the pros and cons of being autistic. There are two identical posts in which the actor talks about how they are consistently making more than an average amount of money by e-Whoring (a social engineering technique applied by actors to imitate partners in cyber-sex and trade the images and videos of the people being imitated on underground forums [47]). They also talk about other money making methods such as buying and selling iPhones. There are three posts in which the actors ask for methods and passwords so they can hack accounts, make money or take revenge. One actor explains how quickly they can crack many games and apps. They claim that their love of exploits and breaking internet communication protocols is due to being autistic, and are obsessed with knowing how things work. The last remaining post is from an actor who claims to be autistic and is interested in "RATing" (a RAT is Remote Access Trojan, malware that allows attackers to gain illegal access to other devices remotely) and "hacking linux".

D. Statistical Analysis Results

Using the methods outlined in §III-G, we produce a sample representing the autistic-claimed population and another representing the non-autistic-claimed population for statistical analysis. From the classified data we have 384 'autistic-claimed' actors. We use the actors' corresponding IDs on the forum to obtain their post counts and post IDs. The post count is referred to as **claimed-post count** hereafter. There are in total 993,002 posts by all autistic-claimed actors. The same query run for the IDs of all 640,074 non-autistic-claimed actors returns 41,119,203 posts (**other-post count**). Figure 5 compares the post distribution for autistic-claimed and non-autistic-claimed populations, using the log-scaled count values to show a clearer "kernel density estimate" (KDE) curves comparison [48].

For the autistic-claimed population, the distribution shows that the density is heavy on the first few thousand posts, meaning many actors in this population seem to be 'chatty' on the forum, with the majority in the 100s and 1000s of posts. On the other hand, the distribution from the non-autistic-claimed population shows that the density skews towards the lower end (10s rather than 100s or 1000s). For this reason, we obtained a matched randomised sample of non-autistic-claimed actors based on post count ranges.

E. Chi-Squared Test of Independence and Results

As outlined in §III-G, Pearson's χ^2 test of independence [46] is used to address our second research question. We carry out two statistical tests using the post volumes as the unit of analysis. We find a significant difference between posts written by the autistic-claimed population and the matched non-autistic claimed sample. Posts written by autistic-claimed actors were significantly less likely to be labelled as criminal in nature ($\chi^2(1, 1, 707, 673) = 7310.3, p < 0.001$). The results are visualised in the mosaic plot in Figure 6. The

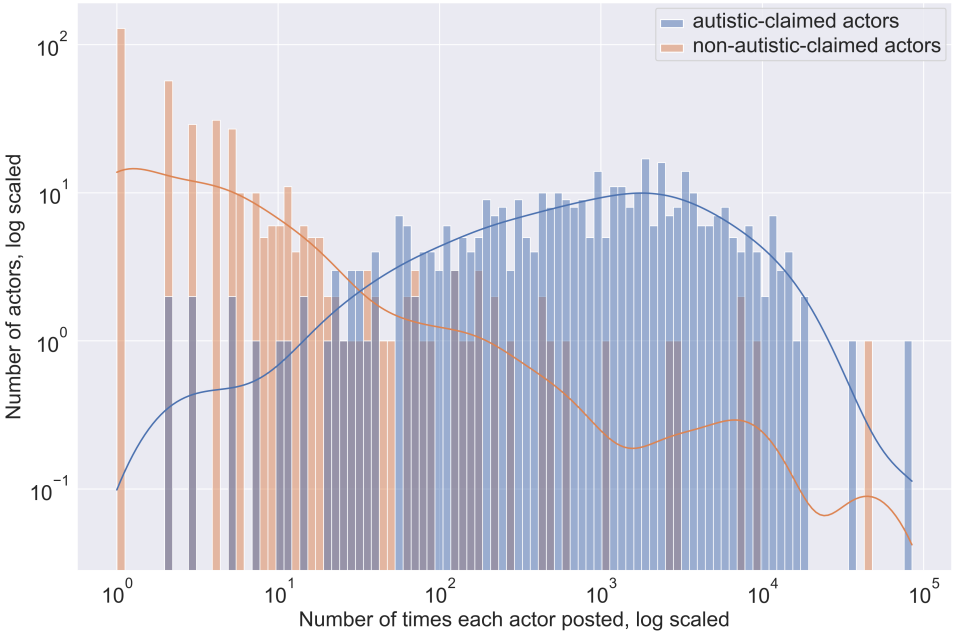


Fig. 5. Distribution of number of posts by autistic-claimed actors vs non-autistic-claimed actors, KDE curve using log scaled counts

standardised residuals show that autistic-claimed actors posted significantly less crime-related content than expected when comparing with the non-autistic-claimed actors, whilst the non-autistic-claimed sample posted significantly fewer posts labelled as non-criminal.

We run a second Table Pearson's χ^2 test of independence on the predicted crime type for each post, excluding the not-criminal class. We find a significant difference in the types of crimes being discussed by the autistic-claimed population and the non-autistic-claimed matched sample ($\chi^2(8, 131, 891) = 1653.04, p < 0.001$). The mosaic plot for this test is shown in Figure 7. The residuals show that autistic-claimed actors post significantly more content related to crime types 'trading credentials', 'VPN/proxy/hosting' and 'access to systems/sql injection' than expected compared with the non-autistic-claimed sample. In contrast, they post significantly less content related to crime types 'DDoS/booting/stress testing' and 'currency exchange' than expected. While they are also more likely to post content related to 'eWhoring', 'identity theft/identity fraud/credit care fraud', and 'spam related/sharing email address/marking' this difference is not significant. Similarly, we find they are less likely to post content related to 'bots/malware' than the non-autistic-claimed actors, but this difference is not significant.

The results of these two χ^2 tests allow us to answer the second research question: **Those who self-disclose to be autistic are less likely to discuss cybercrime compared to a matched sample on the same forum**. This finding is consistent with the research by Payne et al. [10] and Seigfried-Spellar et al. [13]. These two studies found that a diagnosis of autism was associated with a lower risk of indications of cybercrime.

Moreover, the 'Criminal Types' χ^2 test provides more insights into which types of cybercrime autistic-claimed actors are discussing. We find that autistic-claimed actors do post significantly more in relation to 'trading credentials', 'VPN/proxy/hosting' and 'access to systems/sql injection' than the non-austic-claimed matched sample. However, posts related to 'DDoS/booting/stress testing' and 'currency exchange' are significantly less likely to be posted by actors who disclose they are autistic.

V. Conclusion

This research aims to provide more evidence to understand the relationship between cybercrime and autism. Previous studies and empirical research had differing conclusions about the relationship between autistic-like traits and engagement in cybercrime activities, but conclusive evidence has been lacking. While previous research used self-report survey responses, this research takes a different approach, instead we extracting data from the the largest and most active English-language cybercrime forum, to bring us closer to where conversations and interactions between actors about cybercrime take place.

After evaluating different NLP models, we developed two classifiers, one to identify the set of actors who have self-disclosed as autistic, and the second to identify the types of conversations on the forum that involve autism-related terms. Out of the different NLP models evaluated, the combination of word embedding and vectorisation using TF-IDF, XGBoost and hyper-parameter tuning achieved the best performance in both classification tasks. This model was chosen for the automated prediction on the HackForums corpus. It produced a classified dataset that included labels for the types of conversations related to autism, and identified 282 autistic-

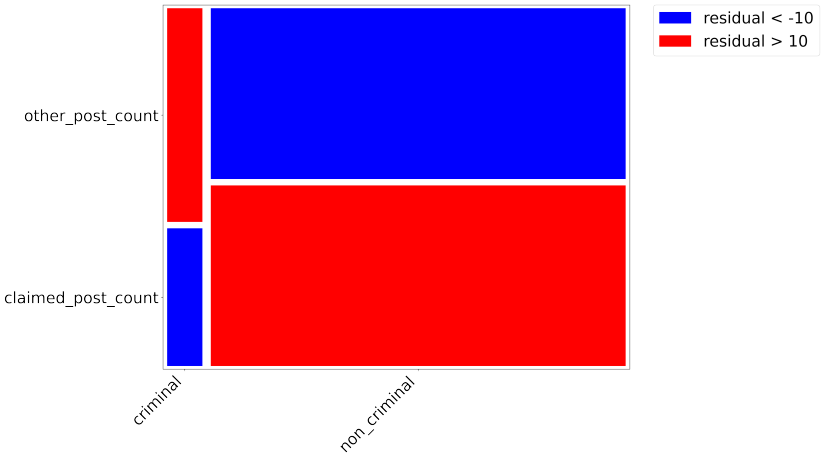


Fig. 6. Mosaic plot for criminal/not criminal posts for autistic-claimed and non-autistic-claimed actors

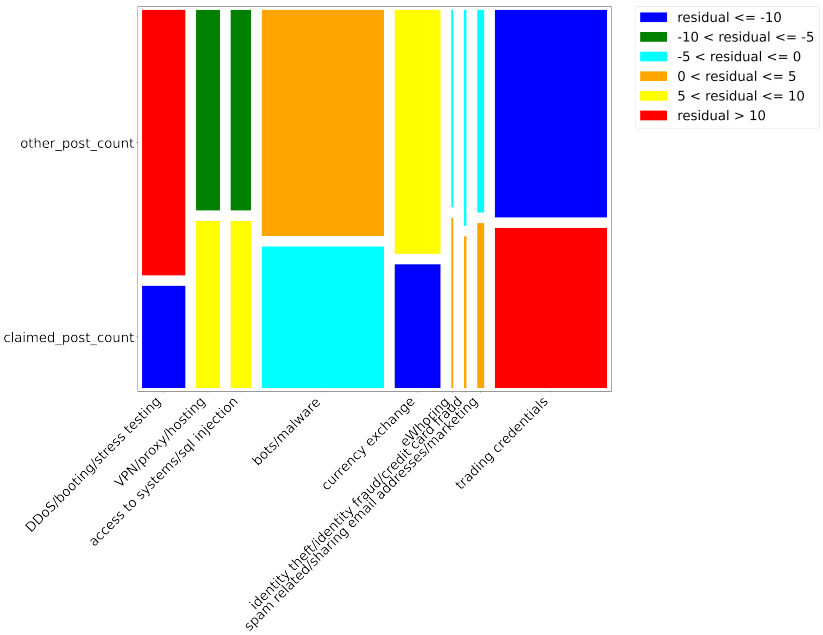


Fig. 7. Mosaic plot for post crime type for autistic-claimed and non-autistic-claimed actors

claimed actors from 19,849 posts. There were false negatives in the non-autistic-claimed set, and so by manually searching for specific autistic-claimed related phrases, a further 102 actors were identified to add to the autistic-claimed population. We extract a matched sample, based on posting volume, from the non-autistic-claimed population.

The topic of discussion arising from the classified dataset suggests that autism-related terms are largely used in a negative context to insult other actors. The terms were rarely used in cybercrime-related context and even less so as a neutralisation for committing cybercrime. The data also showed that non-autistic-claimed actors are much more likely to use the terms for insults whilst the autistic-claimed actors are more likely to use the terms for general-health or neutral conversations.

It appears that, from the classified data, autism-claimed actors are more active on the forum than those that do not reveal they are autistic, meaning they post more frequently than the majority of other actors. This might be partly related to some of the characteristics shared by the National Autistic Society [4], that many autistic people can become experts in their special interests and often like to share their knowledge. As the majority of non-autistic-claimed actors post relatively little, for comparative purposes the sample from this population was extracted based on similar posting volumes as the autistic claimed population. The posts by the two groups of actors were then classified by the crime types as defined by Atondo Siu et al. [16] and compared. The results show that not only are the autistic-claimed actors significantly less likely to be involved in cybercrime-related conversa-

tions, the types of cybercrime discussed are quite different between the two populations. The autistic-claimed actors are significantly more likely to discuss topics related to 'trading credentials', 'VPN/proxy/hosting', and 'access to systems/sql injection', but are significantly less likely to post in related to 'DDos/booting/stress testing', and 'currency exchange'.

In conclusion, actors on HackForums use autism-related terms to mostly insult other actors. Those who self-disclosed as autistic are more likely to participate in non-criminal conversations on the forum. Moreover, the types of crime that they do discuss are different to what the non-autistic-claimed actors are likely to be involved in.

A. Limitations

This research has attempted to overcome the significant difficulties associated with this challenging area of research. However, a number of limitations of the research design were identified. The NLP techniques we apply are not without limitations. The corpus obtained from HackForums posts is full of informal, ill-formed, and non-grammatical text, whilst word embedding and other machine learning methods expect the corpus to be well-formed and grammatical. The predictions by the model are not 100% accurate, hence the autistic-claimed actors used for the analysis contained some false positive values. To minimise the negative impact of this limitation, we use models that had already been tried and tested in past studies. We apply further tuning and the dataset was adjusted appropriately to increase the accuracy. The imbalanced data has also affected the NLP performance, with poor recall values for all models. That means the predictions have a high number of false negatives, which resulted in missing data for the 'claimed' category.

Another limitation is that obtaining ground truth is difficult. The non-autistic-claimed actors may include those who are autistic, but who have not declared this, perhaps due to its negative connotation on the forum. Likewise, some who claim to be autistic may be mistaken, and there was not enough granularity to compare those who disclose they have a diagnosis of autism compared to those that have not been tested. We further note that an autism diagnosis can take many years, and therefore there may be many forum users who remain unaware that they are autistic.

Due to the anonymity of users on HackForums, this study cannot look further into the background, developmental history, and personal characteristics such as age and gender when comparing the self-disclosed autistic actors and the other actors. Such considerations are outside the scope of this research.

B. Future Research

There are a few ways this research could be extended further. First, conversations in HackForums are organised by threads, with one thread per conversation and multiple posts within each thread. There might be more information related to the social relationship and interactions between the different types of actors we could extract if we had included threads. Second, to limit the scope for this work only HackForums was considered. Future research could use multiple underground forums. Third, the autism-related terms used to extract data were selected based on common knowledge. If possible future researchers on related topics could collaborate with autism specialists and law enforcement to improve the accuracy of dataset and analysis. Fourth, the analysis from the classified data indicated that autism-claimed actors posted more frequently, moreover, the types of posts they shared had a different distribution from the non-autism-claimed population. Further investigation could look into motivation and pathways of autistic cybercrime offenders, and build upon the findings from Payne et al. [19] regarding their research on autistic traits and motivations for engaging or declining in cyber-dependent offending. Related to this, one could look into whether controlling for how long an actor is active can change the results. Were the actors in the autistic-claimed population active on the forum a certain time before they started disclosing they were autistic?

Another research direction, using a similar method of applying NLP and machine learning techniques on underground forums, could be to look at the probability of self-disclosed autistic people being victims of cybercrime. Ledingham and Mills [17] stated that autistic individuals have been reported as victims and offenders in computer-related offences such as trolling and fraud. As discovered from this research, autism-related terms were used largely to insult other actors and seldom positively. There were posts that did state the positive traits of being autistic, so what is the likelihood of the autistic traits being exploited or targeted?

Already mentioned in the Limitations section above, future research could evaluate other NLP and machine learning models and look to improve their performance. One of the main issues in our classification was the imbalanced data. This could potentially be improved by a larger annotated sample, a bigger corpus (e.g. include data from more underground forums), and a more refined set of categories.

Whilst this research is a step in the right direction for us to understand more about the behaviour of cyber-criminals and autism, we hope the work produced here provides a solid foundation for future research to answer the many questions which remain.

Acknowledgments

We thank the Cambridge Cybercrime Centre for allowing us access to the CrimeBB dataset, as well as our colleagues for their support and advice. This work is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 949127).

References

[1] C. S. Allely, S. Kennedy, and I. Warren, "Psychiatric and legal issues surrounding the extradition of WikiLeaks founder Julian Assange: The importance of considering the diagnosis of autism spectrum disorder." *Psychology, Public Policy, and Law*, vol. 28, no. 4, pp. 630–643, 2022.

[2] J. Sharp, *Saving Gary McKinnon: A mother's story*. Biteback Publishing, 2013.

[3] G. Davies, "Court of Appeal High Court: Extradition, forum bar and concurrent jurisdiction: Is the case of Love a precedent for trying hackers in the UK? Lauri Love v (1) The Government of the United States of America (2) Liberty [2018] EWHC 172," *The Journal of Criminal Law*, vol. 82, no. 4, pp. 296–300, 2018.

[4] What is autism? Last accessed: 2023-07-24. [Online]. Available: https://www.autism.org.uk/advice-and-guidance/what-is-autism

[5] S. Baron-Cohen, S. Wheelwright, R. Skinner, J. Martin, and E. Clubley, "The Autism-Spectrum Quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians," *Journal of Autism and Developmental Disorders*, vol. 31, no. 1, pp. 5–17, 2001.

[6] L.-O. Lundqvist and H. Lindner, "Is the Autism-Spectrum Quotient a valid measure of traits associated with the autism spectrum? A Rasch validation in adults with and without autism spectrum disorders," *Journal of autism and developmental disorders*, vol. 47, pp. 2080–2091, 2017.

[7] K. L. Ashwood, N. Gillan, J. Horder, H. Hayward, E. Woodhouse, F. S. McEwen, J. Findon, H. Eklund, D. Spain, C. E. Wilson, T. Cadman, S. Young, V. Stoencheva, C. M. Murphy, D. Robertson, T. Charman, P. Bolton, K. Glaser, P. Asherson, E. Simonoff, and M. D. G., "Predicting the diagnosis of autism in adults using the Autism Spectrum Quotient (AQ) questionnaire," *Psychological Medicine*, vol. 46, no. 12, p. 2595–2604, 2016.

[8] S. Baron-Cohen, E. Ashwin, C. Ashwin, T. Tavassoli, and B. Chakrabarti, "Talent in autism: hyper-systemizing, hyper-attention to detail and sensory hypersensitivity," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1522, pp. 1377–1383, 2009.

[9] A. Lim, N. Brewer, and R. L. Young, "Revisiting the relationship between cybercrime, autistic traits, and autism," *Journal of Autism and Developmental Disorders*, pp. 1–12, 2021.

[10] K.-L. Payne, A. Russell, R. Mills, K. Maras, D. Rai, and M. Brosnan, "Is there a relationship between cyber-dependent crime, autistic-like traits and autism?" *Journal of Autism and Developmental Disorders*, vol. 49, no. 10, pp. 4159–4169, 2019.

[11] K. Crawley. The surprising truth about cybersecurity and autism. [Online]. Available: https://cybersecurity.att.com/blogs/security-essentials/the-surprising-truth-about-cybersecurity-and-autism

[12] G. White. Can autism be used as hacking defence? [Online]. Available: https://www.channel4.com/news/can-autism-be-used-as-hacking-defence

[13] K. C. Seigfried-Spellar, C. L. O'Quinn, and K. N. Treadway, "Assessing the relationship between autistic traits and cyberdeviancy in a sample of college students," *Behaviour & Information Technology*, vol. 34, no. 5, pp. 533–542, 2015.

[14] S. Pastrana, A. Hutchings, A. Caines, and P. Buttery, "Characterizing Eve: Analysing cybercrime actors in a large underground forum," in *International Symposium on Research in Attacks, Intrusions, and Defenses*. Springer, 2018, pp. 207–227.

[15] S. Pastrana, D. R. Thomas, A. Hutchings, and R. Clayton, "CrimeBB: Enabling cybercrime research on underground forums at scale," in *Proceedings of the 2018 World Wide Web Conference (WWW '18)*. International World Wide Web Conferences Steering Committee, 2018, p. 1845–1854.

[16] G. A. Siu, B. Collier, and A. Hutchings, "Follow the money: The relationship between currency exchange and illicit behaviour in an underground forum," in *2021 IEEE European Symposium on Security and Privacy Workshops*. IEEE, 2021, pp. 191–201.

[17] R. Ledingham and R. Mills, "A preliminary study of autism and cybercrime in the context of international law enforcement," *Advances in Autism*, 2015.

[18] National Crime Agency. (2017, 1) Pathways Into Cyber Crime. [Online]. Available: https://www.nationalcrimeagency.gov.uk/who-we-are/publications/6-pathways-into-cyber-crime-1/file

[19] K.-L. Payne, K. Maras, R. Mills, A. J. Russell, and M. J. Brosnan, "Self-reported motivations for engaging in or desisting from cyber-dependent offending and the role of autistic traits." *Research in Developmental Disabilities*, vol. 50, 2020.

[20] M. K. Rogers, K. Seigfried, and K. Tidke, "Self-reported computer criminal behavior: A psychological analysis," *Digital Investigation*, vol. 3, pp. 116–120, 2006.

[21] National Crime Agency. (2022, 6) Youth Pathways into Cyber Crime in the UK. [Online]. Available: https://nationalcrimeagency.gov.uk/who-we-are/publications/596-nac-youth-pathways-into-cyber-crime/file

[22] A. Hutchings, "Cybercrime trajectories: An integrated theory of initiation, maintenance, and desistance," *Crime online: Correlates, causes, and context*, pp. 117–140, 2016.

[23] E. H. Sutherland, *White collar crime: The uncut version*. Yale University Press, 1983.

[24] A. V. Vu, J. Hughes, I. Pete, B. Collier, Y. T. Chua, I. Shumailov, and A. Hutchings, "Turning up the dial: the evolution of a cybercrime market through set-up, stable, and covid-19 eras," in *Proceedings of the ACM Internet Measurement Conference*, 2020, pp. 551–566.

[25] J. Franklin, A. Perrig, V. Paxson, and S. Savage, "An inquiry into the nature and causes of the wealth of internet miscreants," in *Proceedings of the 14th ACM Conference on Computer and Communications Security*, ser. CCS '07, 2007, p. 375–388.

[26] A. Caines, S. Pastrana, A. Hutchings, and P. J. Buttery, "Automatically identifying the function and intent of posts in underground forums," *Crime Science*, vol. 7, no. 1, pp. 1–14, 2018.

[27] M. Bada and I. Pete, "An exploration of the cybercrime ecosystem around Shodan," in *2020 7th International Conference on Internet of Things: Systems, management and security (IOTSMS)*. IEEE, 2020, pp. 1–8.

[28] F. Moreno-Vera, M. Nogueira, C. Figueiredo, D. S. Menasché, M. Bicudo, A. Woiwood, E. Lovat, A. Kocheturov, and L. P. de Aguiar, "Cream skimming the underground: Identifying relevant information points from online forums," *IEEE International, Conference on Cyber Security and Resilience (IEEE CSR)*, 2023.

[29] A. A. Paracha, J. Arshad, and M. M. Khan, "Sus you're sus!—identifying influencer hackers on dark web social networks," *Computers and Electrical Engineering*, vol. 107, p. 108627, 2023.

[30] G. M. Sykes and D. Matza, "Techniques of neutralization: A theory of delinquency," *American Sociological Review*, vol. 22, no. 6, pp. 664–670, 1957.

[31] R. G. Morris, "Computer hacking and the techniques of neutralization: An empirical assessment," in *Corporate hacking and technology-driven crime: Social dynamics and implications*. IGI Global, 2011, pp. 1–17.

[32] Y. T. Chua and T. J. Holt, "A cross-national examination of the techniques of neutralization to account for hacking behaviors," *Victims & Offenders*, vol. 11, no. 4, pp. 534–555, 2016.

[33] R. Brewer, S. Fox, and C. Miller, "Applying the techniques of neutralization to the study of cybercrime," *The Palgrave Handbook of International cCybercrime and Cyberdeviance*, pp. 547–565, 2020.

[34] A. Hutchings and R. Clayton, "Exploring the provision of online booter services," *Deviant Behavior*, vol. 37, no. 10, pp. 1163–1178, 2016.

[35] R. S. Portnoff, S. Afroz, G. Durrett, J. K. Kummerfeld, T. Berg-Kirkpatrick, D. McCoy, K. Levchenko, and V. Paxson, "Tools for automated analysis of cybercriminal markets," in *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2017, p. 657–666.

[36] C. C. Centre. (2021) Cambridge Cybercrime Centre: Description of available datasets. [Online]. Available: https://www.cambridgecybercrime.uk/datasets.html

[37] S. B. Kotsiantis, I. Zaharakis, P. Pintelas *et al.*, "Supervised machine learning: A review of classification techniques," *Emerging Artificial Intelligence Applications in Computer Engineering*, vol. 160, no. 1, pp. 3–24, 2007.

[38] J. L. Fleiss, "Measuring nominal scale agreement among many raters." *Psychological Bulletin*, vol. 76, no. 5, p. 378, 1971.

[39] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, pp. 159–174, 1977.

[40] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, pp. 1735–80, 12 1997.

[41] F. Gers and E. Schmidhuber, "LSTM recurrent networks learn simple context-free and context-sensitive languages," *IEEE Transactions on Neural Networks*, vol. 12, no. 6, pp. 1333–1340, 2001.

[42] G. Liu and J. Guo, "Bidirectional lstm with attention mechanism and convolutional layer for text classification," *Neurocomputing*, vol. 337, pp. 325–338, 2019.

[43] D. Jurafsky and J. H. Martin. Vector semantics and embeddings. [Online]. Available: https://web.stanford.edu/~jurafsky/slp3/6.pdf

[44] imbalanced learn.org. (2022, 12) Imbalanced-learn Python API. [Online]. Available: https://imbalanced-learn.org/stable/introduction.html

[45] K. W. Bowyer, N. V. Chawla, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *CoRR*, vol. abs/1106.1813, 2011. [Online]. Available: http://arxiv.org/abs/1106.1813

[46] K. P. F.R.S., "X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 50, no. 302, pp. 157–175, 1900. [Online]. Available: https://doi.org/10.1080/14786440009463897

[47] A. Hutchings and S. Pastrana, "Understanding ewhoring," in *2019 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 2019, pp. 201–214.

[48] M. Conlen. Kernel Density Estimation. [Online]. Available: https://mathisonian.github.io/kde/